DOI:10.11931/guihaia.gxzw201906024

基于叶绿体基因组探讨桃金娘目及其近缘类群的系统发育关系

王雪芹[1]，宋卫武[2*]，肖建加[1]，李超琼[1]，刘红占[1]

（1. 周口师范学院 生命科学与农学学院，河南 周口 466001；2. 周口师范学院 化学化工学院，河南 周口 466001）

摘要：该研究基于叶绿体基因组数据，对桃金娘目（6 科 44 属 97 种）及其近缘类群（牻牛儿苗目 2 科 5 属 25 种）的系统发育关系进行了分析。结果表明：（1）桃金娘目基因组大小为 15.2~17.1 kb，包括的蛋白质编码基因数目为 74~90 个，牻牛儿苗目基因组大小为 11.6~24.2 kb，包括的蛋白质编码基因数目为 75~132 个；（2）对比叶绿体基因组序列和蛋白质编码基因所构建的系统发育树结果，在目间及牻牛儿苗目内差异显著，桃金娘目内基本一致；（3）基于蛋白质编码基因所构建的系统发育树表明，桃金娘目和牻牛儿苗目均为单系，为姐妹类群；桃金娘目内形成两个大支，桃金娘科、Vochysiaceae、野牡丹科形成一支，其中桃金娘科和 Vochysiaceae 关系较近是姐妹群，柳叶菜科、千屈菜科和使君子科形成另一支，其中柳叶菜科和千屈菜科关系较近为姐妹群；科级水平，桃金娘科、Vochysiaceae、野牡丹科、柳叶菜科、千屈菜科、使君子科和牻牛儿苗科均为单系（仅包括一个物种的科除外）；（4）支持将石榴属及菱属置于千屈菜科；（5）对蛋白质编码基因序列变异分析的结果表明，野牡丹科 19 个属的共享变异基因数目为 53 个，变异百分数范围为 5.84%~29.53%，桃金娘科 9 个属的共享变异基因数目为 57 个，其变异百分数范围为 1.31%~15.78%。本研究结果为进一步研究桃金娘目及相关科属的系统发育提供了理论依据。

关键词：牻牛儿苗目，桃金娘科，野牡丹科，蛋白质编码基因，系统发育

中图分类号：Q941　　　　**文献标识码**：A

Phylogeny of Myrtales and related groups based on chloroplast genome

WANG Xueqin[1], SONG Weiwu[2]*, XIAO Jianjia[1], LI Chaoqiong[1], LIU Hongzhan[1]

（1. *College of Life Science and Agronomy*, *Zhoukou Normal University*, Zhoukou 466001, Henan, China; 2. *School of Chemistry & Chemical Engineering*, *Zhoukou Normal University*, Zhoukou 466001, Henan, China）

Abstract: In this study, we used the information on the chloroplast genomes to analyze the phylogenetic relationships of Myrtales (97 species representing 44 genera in six family) and related groups (Geraniales, 25 species representing five genera in two family). The genome size of Myrtales ranged from 15.2 kilo base pairs (kb) to 17.1 kb, including 74 to 90 protein-coding genes. The genome size of Geraniales ranged from 11.6 kb to 24.2 kb, including 75 to 132 protein-coding genes. Phylogenetic analyses of whole genome and protein-coding genes yielded contradicting

基金项目：国家自然科学基金青年基金(81803388)；河南省重点研发与推广专项项目(182102310617)；河南省高等学校重点科研项目(15A180066)[Supported by the National Natural Science Foundation of China (81803388); Key Research Program of Henan Province (182102310617); Key Project of Higher Education of Henan Province (15A180066)]。

作者简介：王雪芹(1986-)，女，湖南邵东人，博士，研究方向为分子系统学，(E-mail) wangxueqin0218@163.com。

***通信作者**：宋卫武，博士，讲师，研究方向为植物资源学，(E-mail) wwsong@zknu.edu.cn。

topologies for intra-order and Geraniales, but congruence results were found in Myrtales. The phylogenetic tree based on protein-coding genes provided strong support for the monophyly of Myrtales and Geraniales and for the placement of Myrtales sister to the Geraniales. Within Myrtales, two major clades were identified. The first clade comprised a Melastomataceae lineage sister to a Myrtaceae + Vochysiaceae lineage and the second clade included Combretaceae sister to a subclade formed by the Onagraceae and Lythraceae lineages. At family level, Myrtaceae, Vochysiacea, Melastomataceae, Lythraceae, Onagraceae, Combretaceae and Geraniaceae were strongly supported as monophyletic (family that represented by only one species was excluded). The placement of *Punica* and *Trapa* in Lythraceae were supported. Additionally, the sequence divergence of the protein-coding genes was estimated. For Melastomataceae, 53 variable protein-coding genes were identified, with the variation percentage ranged from 5.84% to 29.53% among the 19 genera. In Myrtaceae, the proportion of variability of 57 variable protein-coding genes ranged from 1.31% to 15.78% among the 9 genera. Out study provided an important framework for further phylogenetic study in Myrtales and related groups.
Key words: Geraniales, Myrtaceae, Melastomataceae, protein-coding genes, phylogeny

在分子系统学发展之初，人们常利用单个基因或者几个基因来解决类群的系统发育关系，对于某些类群而言，可能得到较理想的结果，而对于某些特殊类群，如进化速率较慢的类群或辐射进化的类群，由于缺乏足够的信息位点，研究遇到阻滞。通过增加基因数目及联合分析的方法，在一定程度上，研究者可以提高系统发育树的分辨率及支持率。随着二代测序技术的发展，测序成本的降低，使得系统发育基因组学应运而生。叶绿体基因组由于其大小适中，便于测序及后期组装拼接，且在不同类群之间具有良好的共线性(张韵洁和李德铢，2011)，因而在系统发育分析中得到广泛应用。截止 2019 年 4 月 NCBI 数据库中的叶绿体基因组数据达 3000 多个，而在 2010 年 10 月仅为 146 个，其数据增长呈爆发之式，为系统发育研究提供了便利及机遇。目前，叶绿体基因组已应用于研究被子植物不同阶层的系统发育关系，如柑橘属(*Citrus*)（Carbonell-Caballero et al.，2015）、棉属(*Gossypium*)（Wu et al.，2018）、禾本科(Poaceae)（Pessoa-Filho et al.，2018）、兰科(Orchidaceae)（Dong et al.，2018）、石竹目 (Caryophyllales) （Liu et al.，2018）、姜目(Zingiberales)（Wu et al.，2018）及被子植物框架（Bausher et al.，2006）。

桃金娘目 Myrtales 隶属于蔷薇支，其为单系类群，得到了形态学和分子系统学证据的支持，而桃金娘目在蔷薇类中的系统位置未定，一直为研究者所关注。基于 *rbcL* 的分析结果表明，桃金娘目在蔷薇类中位置不确定（Hilu et al.，2003）。Zhu et al.（2007）基于 1 个线粒体基因、2 个叶绿体及 1 个核基因片段联合分析的结果表明，桃金娘目可能和牻牛儿苗目(Geraniales)的关系较近（无支持率），共同组成了蔷薇类的基部类群。而 Wang et al.（2009）基于 12 个基因（10 个叶绿体基因、2 个核基因）联合分析的结果表明，桃金娘目和牻牛儿苗目关系较近，形成中等支持率的一支，均隶属于锦葵类植物 malvids（即真蔷薇二类植物 eurosids II），为锦葵类植物其余科的姐妹群。

桃金娘目包含 9 科 380 属，约 13 000 种（The Angiosperm Phylogeny Group，2016），囊括了 6%左右的核心真双子叶植物类群。目下所包括的 9 个科中，Crypteroniaceae、Alzateaceae 及 Penaeaceae 所包括的物种数较少，99%以上的物种分布于其余 6 科，依次为桃金娘科(Myrtaceae)、野牡丹科(Melastomataceae)、柳叶菜科(Onagraceae)、千屈菜科(Lythraceae)、使君子科(Combretaceae)及 Vochysiaceae。关于桃金娘目的分子系统学研究主要有：Conti et al.（1997）基于 *rbcL* 序列的研究结果表明桃金娘科可能与野牡丹科关系较近，柳叶菜科和千屈菜科是姐妹群，而使君子科系统位置未定。Sytsma et al.（2004）基于 *rbcL* 和 *ndhF* 序列，

进一步扩大采样范围，研究了桃金娘目内科之间的系统发育关系，结论表明使君子科可能是该目的基部类群，与目内其它科为姐妹群关系（支持率不高），桃金娘科可能是野牡丹科+Memecylaceae 的姐妹群，柳叶菜科和千屈菜科为姐妹群关系获较高支持率。Soltis et al.（2011）基于 17 个基因的数据结果表明桃金娘目和牻牛儿苗目为姐妹群，获中等支持率支持。

综上，桃金娘目的近缘类群未定，目下科之间的关系尚需进一步研究。已有的分子系统学研究表明在该类群中，基于单个基因或几个基因的联合分析，均不能得到稳定的系统发育树。本研究利用公共数据库中已公开发表的桃金娘目 6 个主要科及其近缘类群的叶绿体基因组数据，通过系统发育树重建，获得高支持率且高分辨率的桃金娘目的系统发育框架，确定了桃金娘目与牻牛儿苗目的姐妹群关系，并进一步确认了使君子科是柳叶菜科和千屈菜科的姐妹群。此外，我们对桃金娘科和野牡丹科的蛋白质编码基因的序列变异程度进行了统计分析，为后续相关科的系统发育研究提供数据参考。

1 材料和方法

1.1 数据收集

从 GenBank 数据库中下载已发表的桃金娘目 6 科 44 属 97 种和牻牛儿苗目 2 科 5 属 25 种的叶绿体全基因组序列及蛋白质编码基因（表 1）。选择锦葵目(Malvales)3 种、十字花目(Brassicales) 2 种、十齿花目(Huerteales) 1 种，无患子目(Sapindales) 5 种，作为外类群（表 1）。利用 Excel 统计每个叶绿体基因组的大小及蛋白质编码基因的数目。

表 1 物种名称，序列编号，基因组大小和基因数目

Table1 Taxa included in the study with GenBank accession numbers, genome size, and number of genes

种名 Taxon	GenBank 编号 GenBank Accession Numbers	序列长度 Length (bp)	蛋白质编码基因数目 Number of protein-coding genes
桃金娘目 Myrtales			
使君子科 Combretaceae			
红榄李 *Lumnitzera littorea*	MG182696.1	159 687	84
Laguncularia racemosa	MK726017.1	160 378	84
Terminalia guyanensis	MK726027.1	159 750	83
千屈菜科 Lythraceae			
Lagerstroemia fauriei	KT358807.1	152 440	85
L. floribunda	KX765488.1	152 240	85
桂林紫薇 *L. guilinensis*	KU885923.1	152 193	85
紫薇 *L. indica*	KF572028.1	152 231	84
紫薇 *L. indica*	KX263727.1	152 205	86
云南紫薇 *L. intermedia*	KX852427.1	152 330	85
L. speciosa	KU821692.1	152 476	85
L. speciosa	KX572149.1	152 526	84
南紫薇 *L. subcostata*	KF572029.1	152 049	84
石榴 *Punica granatum*	KY635883.1	158 633	84
菱科 Trapaceae			
Trapa maximowiczii	KY705084.1	155 577	82

野牡丹科 Melastomataceae			
Allomaieta villosa	KX826819.1	156 452	85
Barthea barthei	KY873324.1	155 951	82
Bertolonia acuminata	KX826820.1	156 045	85
Blakea schlimii	KX826821.1	155 862	85
Eriocnema fulva	KX826822.1	155 994	85
Graffenrieda moritziana	KX826823.1	155 733	84
Henriettea barkeri	KX826824.1	156 527	85
Melastoma candidum	KY745894.1	156 682	85
Merianthera pulchra	KX826825.1	156 168	85
Miconia dodecandra	KX826826.1	157 216	85
Nepsera aquatica	KX826827.1	155 110	85
Opisthocentra clidemioides	KX826828.1	156 352	85
Pterogastra divaricata	KX826829.1	154 948	84
Rhexia virginica	KX826830.1	154 635	85
Rhynchanthera bracteata	KX826831.1	155 108	85
Salpinga maranonensis	KX826832.1	153 311	83
Tibouchina longifolia	KX826833.1	156 789	85
虎颜花 *Tigridiopalma magnifica*	MF663760.1	155 663	84
Triolena amazonica	KX826834.1	156 652	85
桃金娘科 Myrtaceae			
Acca sellowiana	KX289887.1	159 370	85
Allosyncarpia ternata	KC180806.1	159 593	85
Angophora costata	KC180805.1	160 326	85
A. floribunda	KC180804.1	160 245	85
Corymbia citriodora subsp. *citriodora*	KP015029.1	159 995	85
C. citriodora subsp. *variegata*	KM885985.1	160 146	84
C. eximia	KC180802.1	160 012	85
C. gummifera	KC180800.1	160 713	85
C. henryi	KP015032.1	160 095	85
C. maculata	KC180801.1	160 045	85
C. tessellaris	KC180803.1	160 127	85
C. torelliana	KP015033.1	159 994	84
Eucalyptus aromaphloia	KC180789.1	160 149	85
E. baxteri	KC180773.1	160 032	85
赤桉 *E. camaldulensis*	KC180791.1	160 164	85
E. cladocalyx	KC180786.1	160 213	85
E. cloeziana	KC180779.1	160 015	85
E. curtisii	KC180782.1	160 038	85
E. deglupta	KC180792.1	160 177	85
E. delegatensis	KC180771.1	159 724	85
E. diversicolor	KC180795.1	160 214	85
E. diversifolia	KC180774.1	159 954	85
E. elata	KC180776.1	159 899	85

E. erythrocorys	KC180799.1	159 742	85
蓝桉 *E. globulus*	KC180787.1	160 267	85
蓝桉 *E. globulus* subsp. *globulus*	AY780259.1	160 286	90
大桉 *E. grandis*	HM347959.1	160 137	74
E. guilfoylei	KC180798.1	160 520	85
E. marginata	KC180781.1	160 076	85
蜜味桉 *E. melliodora*	KC180784.1	160 386	85
小帽桉 *E. microcorys*	KC180797.1	160 225	85
E. nitens	KC180788.1	160 271	85
E. obliqua	KC180769.1	159 527	85
E. patens	KC180780.1	160 187	85
E. pauciflora	MG921592.1	159 942	84
E. polybractea	KC180785.1	160 268	85
E. radiata	KC180770.1	159 529	85
E. regnans	KC180777.1	160 031	85
柳叶桉 *E. saligna*	KC180790.1	160 015	85
E. salmonophloia	KC180796.1	160 413	85
E. sieberi	KC180775.1	159 985	85
E. spathulata	KC180793.1	161 071	85
E. torquata	KC180794.1	160 223	85
E. umbra	KC180778.1	159 576	85
E. verrucata	KC180772.1	160 109	85
Plinia trunciflora	KU318111.1	159 512	84
番石榴 *Psidium guajava*	KX364403.1	158 841	84
番石榴 *P. guajava*	KY635879.1	158 896	83
Stockwellia quadrifida	KC180807.1	159 561	85
乌墨 *Syzygium cumini*	GQ870669.3	160 373	87
柳叶菜科 Onagraceae			
Epilobium ulleungensis	MH198310.1	160 912	85
毛草龙 *Ludwigia octovalvis*	KX827312.1	159 396	82
Oenothera argillicola	EU262887.2	165 061	84
月见草 *O. biennis*	EU262889.2	164 796	84
O. elata subsp. *elata*	KT881169.2	165 403	84
O. elata subsp. *hookeri*	KT881170.1	165 359	84
黄花月见草 *O. glazioviana*	EU262890.2	165 359	84
O. grandiflora	KT881173.1	166 545	84
曲序月见草 *O. oakesiana*	KT881176.1	163 575	84
小花月见草 *O. parviflora*	EU262891.2	163 367	84
O. picensis subsp. *picensis*	KX118607.1	167 092	87
O. villaricae	KX118606.1	165 779	87
长毛月见草 *O. villosa* subsp. *villosa*	KX687910.1	164 312	84
Vochysiaceae			
Vochysia acuminata	MK726031.1	171 315	87
Salvertia convallariodora	MK726026.1	171 267	87

Ruizterania albiflora	MK726023.1	162 345	84
Qualea grandiflora	MK726022.1	163 026	84
Korupodendron songweanum	MK726013.1	161 149	83
Erisma bracteosum	MK726009.1	160 687	85
Callisthene erythroclada	MK726008.1	161 626	84
牻牛儿苗目 Geraniales			
牻牛儿苗科 Geraniaceae			
Erodium absinthoides	KJ523886.1	162 618	90
E. carvifolium	HQ713469.1	116 935	75
E. chrysanthum	KJ701602.1	168 946	97
Monsonia emarginata	KT692738.1	156 877	82
M. marlothii	KT692739.1	134 416	75
Pelargonium alternans	KF240617.1	173 374	97
P. australe	KM459517.1	165 508	97
P. citronellum	KM527888.1	171 223	97
P. cotyledonis	KM459516.1	166 111	97
P. cucullatum	KM527887.1	170 963	97
P. dichondrifolium	KM459515.1	167 836	97
P. dolomiticum	KM527889.1	202 749	117
P. echinatum	KM527891.1	170 212	97
P. exhibens	KM527890.1	171 665	98
P. exstipulatum	KM527892.1	168 732	97
P. fulgidum	KM527893.1	171 502	97
P. incrassatum	KM527894.1	173 196	97
P. myrrhifolium	KM527895.1	173 272	99
P. nanum	KM527896.1	173 496	97
P. quercifolium	KM527897.1	170 569	97
P. tetragonum	KM527899.1	173 410	99
P. transvaalense	KM527900.1	242 575	132
P. trifidum	KM527898.1	199 553	116
Hypseocharis bilobata	KF240616.1	165 002	85
Vivianiaceae			
Viviania marifolia	KF240615.1	157 291	86
外类群 Outgroups			
十字花目 Brassicales			
十字花科 Brassicaceae			
鼠耳芥 *Arabidopsis thaliana*	KX551970.1	154 515	83
欧洲油菜 *Brassica napus*	GQ861354.1	152 860	87
十齿花目 Huerteales			
瘿椒树科 Tapisciaceae			
瘿椒树 *Tapiscia sinensis*	MF926267.1	161 100	85
锦葵目 Malvales			
梧桐科 Sterculiaceae			
可可 *Theobroma cacao*	HQ336404.2	160 604	85

锦葵科 Malvaceae			
陆地棉 *Gossypium hirsutum*	DQ345959.1	160 301	85
龙脑香科 Dipterocarpaceae			
Shorea pachyphylla	MH841940.1	150 753	81
无患子目 Sapindales			
槭树科 Aceraceae			
三角枫 *Acer buergerianum*	KY419137.1	156 461	88
芸香科 Rutaceae			
甜橙 *Citrus sinensis*	DQ864733.1	160 129	89
橄榄科 Burseraceae			
Commiphora wightii	MF957201.1	156 064	86
苦木科 Simaroubaceae			
Leitneria floridana	KT692940.1	158 763	87
漆树科 Anacardiaceae			
芒果 *Mangifera indica*	KY635882.1	157 780	83

1.2 构建矩阵

全基因组序列使用 CIPRES Science Gateway（Miller et al.，2010）在线工具 MAFFT v 7.394（Katoh & Standley，2013）软件对序列进行比对，参数设置为默认值。每个物种的蛋白质编码基因按字母排序，连接，采用 CIPRES Science Gateway 在线工具 MAFFT v 7.394 软件对序列进行比对。比对结果在 MEGA 7.0 软件（Kumar et al.，2016）中进行部分人工校正。

1.3 系统发育分析

基于所获得的矩阵分别采用最大似然法（Maximum likelihood, ML）及贝叶斯推断（BI）进行系统发育分析。RaxML 分析运用软件 RaxML ver.8.2.8（Stamatakis，2014），选择 GTRGAMMA 模型，采用快速靴代值（rapid bootstrap）分析，重复 1000 次（Bootstrap, BS）。应用 MrBayes v 3.2.5（Ronquist & Huelsenbeck，2003）进行贝叶斯推断分析，选择 GTR+GAMMA 模型。BI 分析的参数设置为：采用 MCMC（Markov chain monte carlo）算法，运行 1 000 000 代，每 1 000 代取样一次，开始的 25%样品作为老化样本（Burn-in samples）舍弃，以剩余样本构建主要规则一致树并计算各分支的后验概率（posterior probability, PP）。

2 结果与分析

2.1 基因组大小

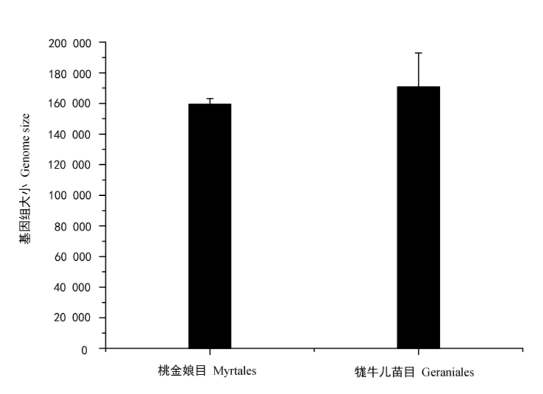
桃金娘目叶绿体全基因组大小在 152 049~171 315 bp 之间，平均值为 15.9 kb，蛋白质编码基因数目在 74~90 之间，平均值为 85（表 2、图 1）。牻牛儿苗目叶绿体全基因组大小在 116 935~242 575 bp 之间，基因组大小平均值为 17.0 kb，蛋白质编码基因数目在 75~132 之间，平均值为 97（表 2，图 1）。

表 2 基因组大小和蛋白质编码基因的比较

Table 2 Comparison of genome size and protein-coding genes

类群 Groups	基因组大小 Genome size		蛋白质编码基因 Protein-coding gene	
	平均值 Mean	标准差 Standard deviation	平均值 Mean	标准差 Standard deviation
桃金娘目 Myrtales	159 374	3 763	85	1

牻牛儿苗目 Geraniales	170 721	22 175	97	12



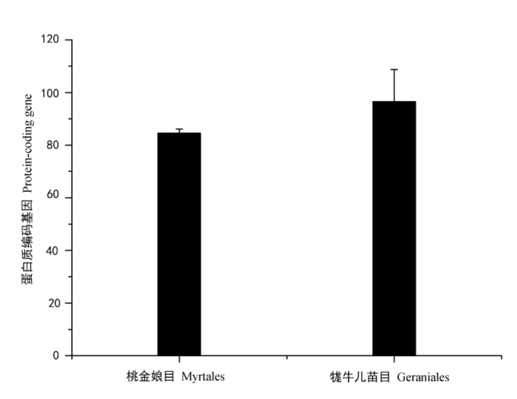


图 1 基因组大小和蛋白质编码基因的比较

Fig. 1 Comparison of genome size and protein-coding genes

2.2 矩阵分析

为了验证叶绿体基因组在不同分类单元系统发育分析中的应用性，我们构建了两套矩阵，即基于叶绿体全基因组序列形成的矩阵和基于蛋白质编码基因形成的矩阵。经序列比对及人工校正我们一共获得 6 个矩阵：（1）由叶绿体全基因组序列构成的 3 个矩阵：由桃金娘目、牻牛儿苗目及外类群构成的包括 139 条序列的矩阵（矩阵 1）、由桃金娘目构成的包括 103 条序列的矩阵（矩阵 3）、由牻牛儿苗目构成的包括 25 条序列的矩阵（矩阵 5）；（2）由蛋白质编码基因构成的 3 个矩阵：相应的由桃金娘目、牻牛儿苗目及外类群构成（矩阵 2）、仅包括桃金娘目（矩阵 4）及仅包括牻牛儿苗目（矩阵 6）的矩阵。

矩阵信息详见表 3。其中由牻牛儿苗目的叶绿体全基因组序列构成的矩阵即矩阵 5 最大，长度为 762 293 bp，其中变异位点的数目为 205 260，信息位点的数目为 93 186（12.22%），缺失数据比例为 77.60%。仅由桃金娘目的蛋白质编码基因构成的矩阵即矩阵 4 最小，长度为 113 097 bp，其中变异位点的数目为 26 162，信息位点的数目为 18 171（16.07%），缺失数据比例为 30.09% （表 3）。

表 3 矩阵信息统计

Table 3 Information of matrix

矩阵 Matrix	矩阵长度 All sites	变异位点 Variable sites	信息位点 Parsimony-informative sites	缺失数据 Missing data
矩阵 1 Matrix 1	592 879	249 452	149 806	72.80%
矩阵 2 Matrix 2	186 769	53 589	39 804	57.36%
矩阵 3 Matrix 3	240 022	87 025	64 590	33.60%
矩阵 4 Matrix 4	113 097	26 162	18 171	30.09%
矩阵 5 Matrix 5	762 293	205 260	93 186	77.60%
矩阵 6 Matrix 6	159 381	31 722	18 004	47.86%

2.3 基于不同矩阵构建系统进化树

我们对基于不同数据集所构建的系统发育树结果进行了比较分析。支持率方面，基于叶绿体全基因组序列与基于蛋白质编码基因的系统发育树在目间，及桃金娘目内没有明显差异。但在牻牛儿苗目内，基于叶绿体全基因组序列构建的系统发育树支持率明显降低，其中支持率≥95%的分支为 68%，而基于蛋白质编码基因所构建的系统发育树中，支持率≥95%的分支为 95% （图 2）。

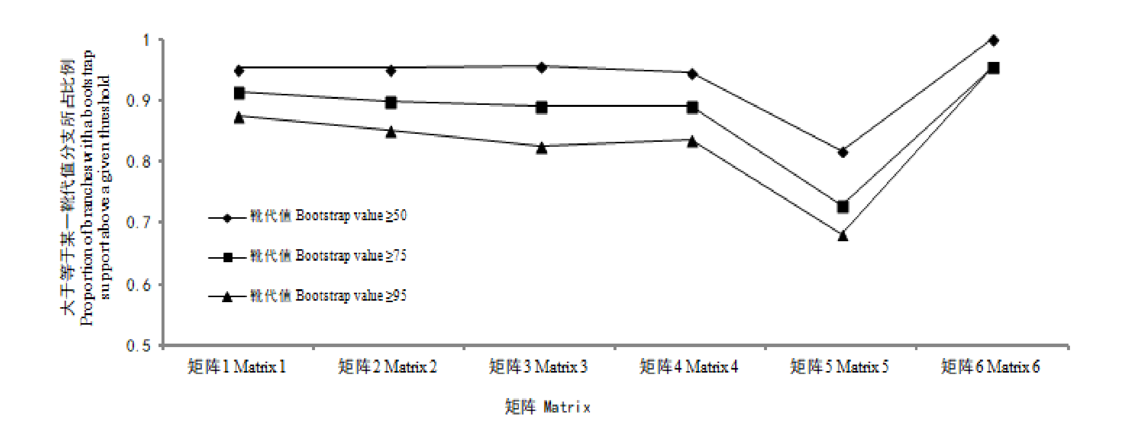
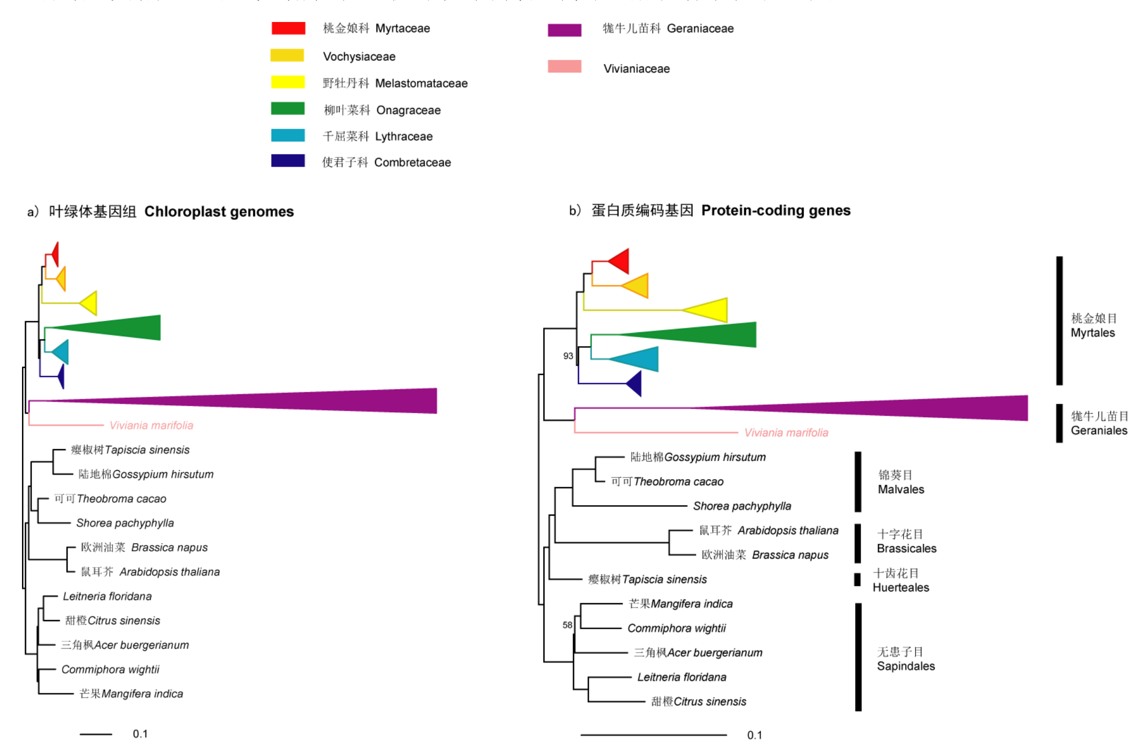


图 2 最大似然法构建系统发育树支持率统计

Fig. 2 Summary of bootstrap values for maximum likelihood analysis

拓扑结构方面。对包括 139 条序列的系统发育树进行分析，两个数据集当中桃金娘目和牻牛儿苗目均形成了单系，支持率为 100%，外类群中的锦葵目在全基因组序列数据集中并未形成单系（图 3）。桃金娘目两个数据集的系统发育树的拓扑结构基本一致，仅在桉属 *Eucalyptus* 和月见草属 *Oenothera* 内部有微小差异。牻牛儿苗目两个数据集所构建系统发育树的拓扑结构则存在显著差异，叶绿体全基因组数据集中，牻牛儿苗科 Geraniaceae 未形成单系，牻牛儿苗科内牻牛儿苗属 *Erodium*、天竺葵属 *Pelargonium* 和 *Monsonia* 均未形成单系，而蛋白质编码基因数据集中，牻牛儿苗科及科下的属均为单系（图 4）。

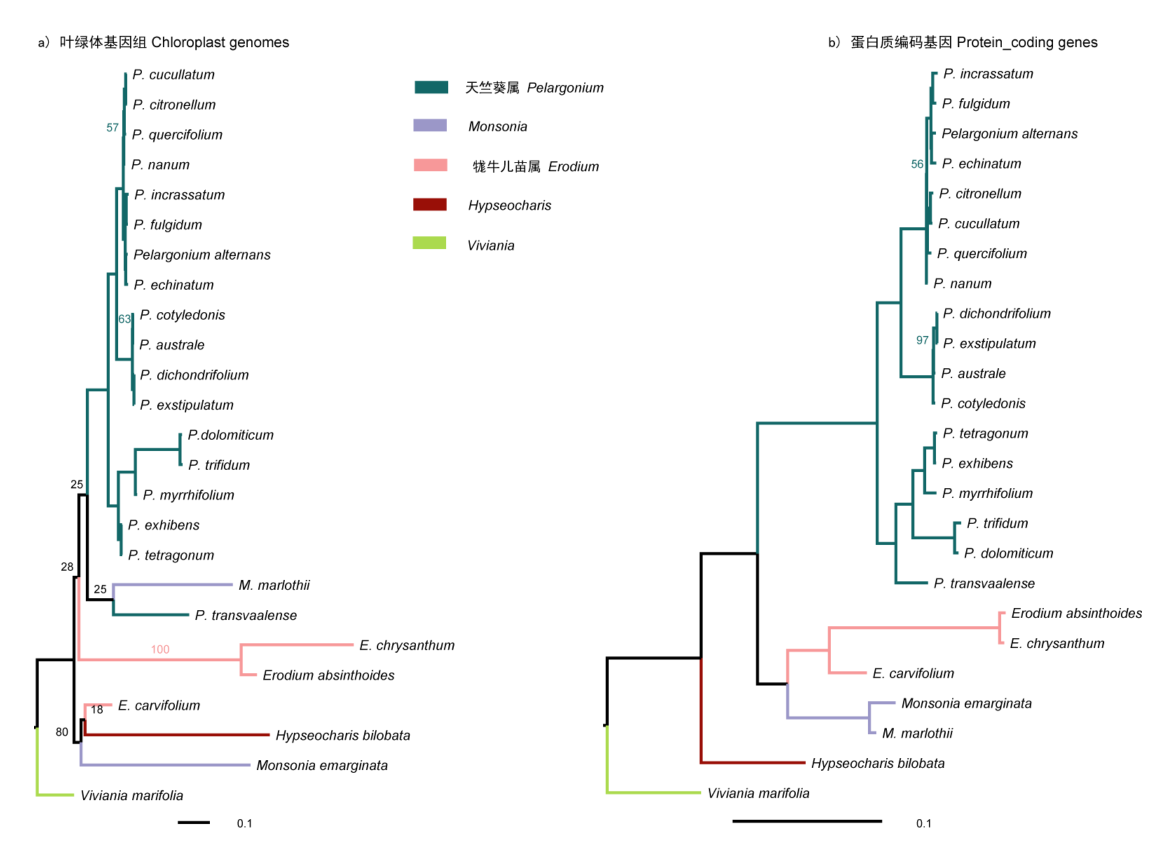


注：数字代表分支的靴代值 BS；靴代值为 100 的不显示。

Note: Nodes are labeled with maximum likelihood bootstrap support; Nodes with 100% bootstrap support in the maximum likelihood analysis are unlabeled.

图 3 基于不同矩阵构建的桃金娘目和牻牛儿苗目的最大似然（ML）树

Fig. 3 Maximum likelihood of phylogenetic hypothesis of Myrtales and Geraniales based different datasets



注：数字代表分支的靴代值 BS；靴代值为 100 的不显示。

Note: Nodes are labeled with maximum likelihood bootstrap support; Nodes with 100% bootstrap support in the maximum likelihood analysis are unlabeled.

图 4 基于不同矩阵构建的牻牛儿苗目的最大似然（ML）树

Fig. 4 Maximum likelihood of phylogenetic hypothesis of Geraniales based different datasets

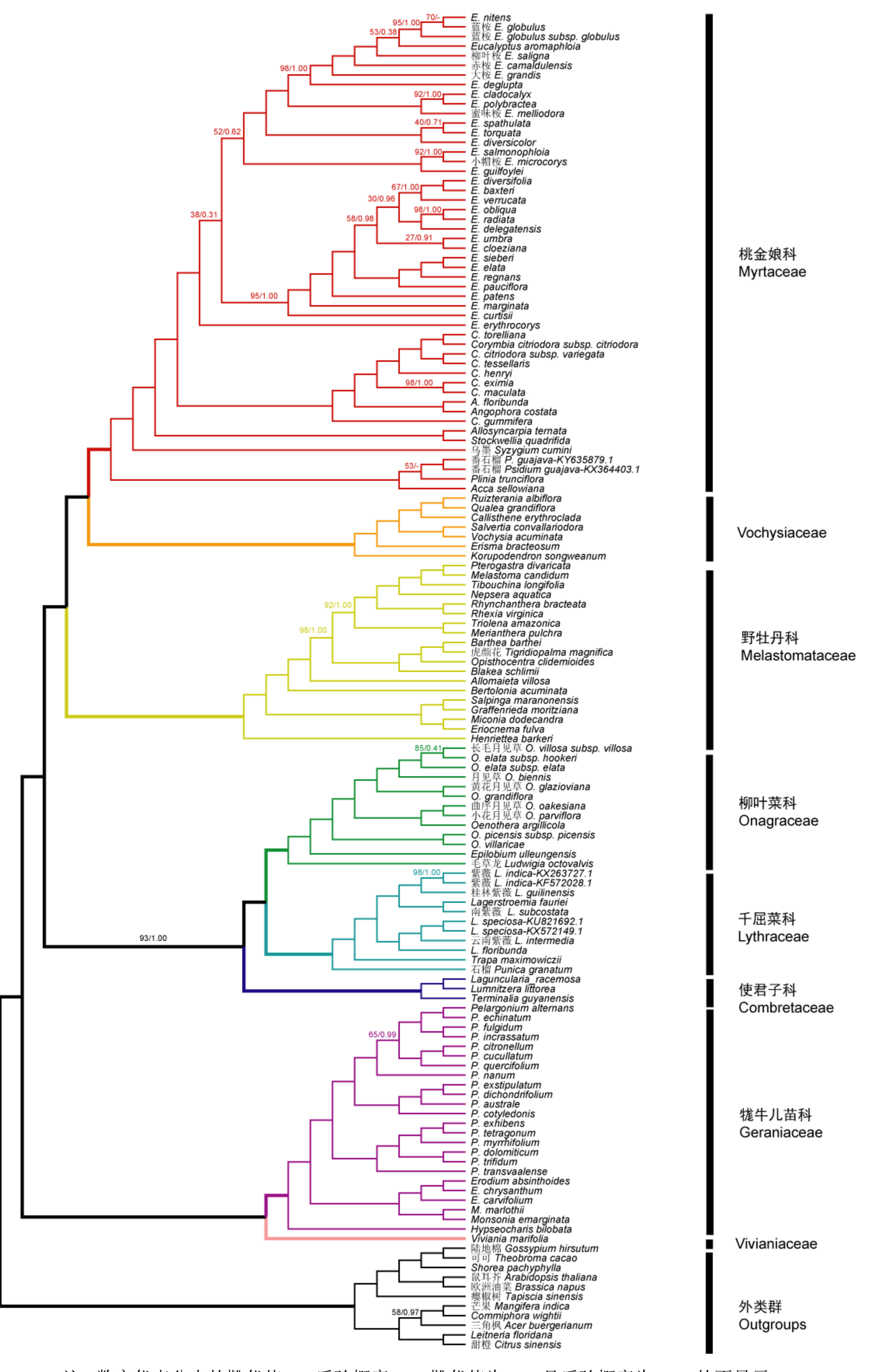
2.4 桃金娘目和牻牛儿苗目的系统发育关系

根据上述结果，我们选择了包括 139 条序列 136 个物种（包括亚种）的蛋白质编码基因的矩阵（矩阵 2），分别进行最大似然法建树及贝叶斯分析。矩阵长度为 186 769 bp，其中变异位点的数目为 53 589，信息位点的数目为 39 804（21.31%），缺失数据比例为 57.36%。采用最大似然法我们获得了高分辨率的系统发育树（图 5），其平均支持率为 95%，支持率≥95%的支所占比重为 87.5%。目级水平，外类群的锦葵目、十字花目、十齿花目、无患子目与内类群的桃金娘目、牻牛儿苗目均为单系（BS=100%；PP=1.0）。科级水平，桃金娘目下的桃金娘科、野牡丹科、柳叶菜科、千屈菜科、使君子科和 Vochysiacea 及牻牛儿苗目下的牻牛儿苗科均为单系（仅包括一个种的科除外）（BS=100%；PP=1.0）。属级水平，桃金娘科的桉属、杯果木属(*Angophora*)，柳叶菜科的月见草属，千屈菜科的紫薇属(*Lagerstroemia*)，牻牛儿苗科的牻牛儿苗属、天竺葵属和 *Monsonia* 均为单系（属下仅包括一个种的除外）（BS=100%；PP=1.0）。桃金娘科的 *Corymbia* 不是单系类群。

桃金娘目为单系类群，所包括 6 个科的科间关系得到解决。桃金娘科和 Vochysiacea 关系较近，形成一支，与野牡丹科是姐妹群。柳叶菜科和千屈菜科是姐妹群，使君子科是柳叶菜科+千屈菜科的姐妹群。千屈菜科这一分支包括了紫薇属、石榴属(*Punica*)及菱属(*Trapa*)，其中石榴属在中国植物志中隶属于石榴科(Punicaceae)，在 *Flora of China* 中将石榴属移置于千屈菜科，而菱属则隶属于菱科(Trapaceae)。柳叶菜科这一分支包括了月见草属、丁香蓼属

(*Ludwigia*)、柳叶菜属(*Epilobium*)，其中丁香蓼属为基部类群，月见草属和柳叶菜属组成一支。野牡丹科包括 19 属，属间的系统关系与前人基于形态学数据（Renner，1993）及叶绿体片段（Renner，2004；Goldenberg et al.，2012）的研究结果有出入。鉴于所包括的属有限，仅占全部属的 10%左右，且每个属仅包括一个物种，野牡丹科的系统发育关系有待进一步扩大样品进行研究。桃金娘科包括桉属(*Eucalyptus*)、*Corymbia*、杯果木属、*Allosyncarpia*、*Stockwelli*、蒲桃属(*Syzygium*)、番石榴属(*Psidium*)、*Acca*、*Plinia* 共 9 属。前人基于 *matK* 的系统发育树结果表明，桃金娘科可以分为两个亚科，17 个族（Wilson et al.，2005）。本研究中所包括的属均隶属于桃金娘亚科(Myrtoideae)，分布于 3 个族，其中蒲桃属隶属于 Syzygieae，番石榴属、*Acca* 和 *Plinia* 隶属于 Myrteae，杯果木属、桉属、*Allosyncarpia*、*Corymbia* 和 *Stockwelli* 隶属于 Eucalypteae，3 个族均为单系（BS=100%；PP=1.0）。

牻牛儿苗目为单系类群，Vivianiaceae 为基部类群，牻牛儿苗科形成了一个单系，其中 *Hypseocharis* 为其它属的姐妹群，牻牛儿苗属和 *Monsonia* 为姐妹群。



注：数字代表分支的靴代值 BS/后验概率 PP；靴代值为 100 且后验概率为 1.00 的不显示。

Note: Nodes are labeled with maximum likelihood bootstrap support (first number) and Bayesian posterior

probability (second number); Nodes with 100% bootstrap support in the maximum likelihood analysis and 1.00 posterior probability in the bayesian analysis are unlabeled.

图 5 基于蛋白质编码基因构建的桃金娘目及其近缘类群的系统发育树

Fig. 5 Topologies of Myrtales and related groups based the protein-coding genes

2.5 蛋白质编码基因的序列分化

为了评估不同蛋白质编码基因在科下系统发育分析中的应用,我们对野牡丹科及桃金娘科序列长度大于 200 bp 的蛋白质编码基因的序列分化程度进行了统计（其余科因为所包括的属较少不在分析之列）。野牡丹科 19 个属的共享变异基因数目为 53 个，变异百分数范围为 5.84%~29.53%，变异率最高的 5 个基因依次为 *rpl22*、*rps15*、*matK*、*ndhF* 和 *rps3*，其中发生插入缺失的基因数目为 12 个(表 4);而桃金娘科 9 个属的共享变异基因数目为 57 个，由于分析所包括的属较野牡丹科而言要少，其变异百分数较低，范围为 1.31% - 15.78%，变异率最高的 5 个基因依次为 *rps12*、*ycf1*、*rpl22*、*rps15* 和 *matK*，发生插入缺失的基因数目为 11 个（表 4）。

表 4 野牡丹科 19 属及桃金娘科 9 属的蛋白质编码基因序列变异统计

Table 4 Sequence divergence of variable coding regions (>200 bp) from 19 chloroplast genomes of Melastomataceae and 9 chloroplast genomes of Myrtaceae

	片段 Fragment	野牡丹科 Melastomataceae						桃金娘科 Myrtaceae					
		比对长度 Aligned length (bp)	变异位点 Variable positions	碱基替换 Nucleotide substitutions	插入缺失数目 Number of indels	插入缺失总长度 Total length of indels	变异百分数 Percent variability (%)	比对长度 Aligned length (bp)	变异位点 Variable positions	碱基替换 Nucleotide substitutions	插入缺失数目 Number of indels	插入缺失总长度 Total length of indels	变异百分数 Percent variability (%)
1	*accD*	1 650	350	270	11	80	21.21	1 479	62	62	0	0	4.19
2	*atpA*	1 512	168	168	0	0	11.11	1 524	37	37	0	0	2.43
3	*atpB*	1 497	133	133	0	0	8.88	1 497	38	38	0	0	2.54
4	*atpE*	423	57	57	0	0	13.48	402	7	7	0	0	1.74
5	*atpF*	555	91	86	1	5	16.40	570	23	23	0	0	4.04
6	*atpH*	246	24	24	0	0	9.76	246	5	5	0	0	2.03
7	*atpI*	744	75	75	0	0	10.08	750	19	19	0	0	2.53
8	*ccsA*	960	204	201	2	3	21.25	960	60	60	0	0	6.25
9	*cemA*	699	109	105	4	4	15.59	690	22	22	0	0	3.19
10	*clpP*	591	71	71	0	0	12.01	591	13	11	1	2	2.20
11	*matK*	**1 572**	**426**	**419**	**4**	**7**	**27.10**	**1 527**	**124**	**124**	**0**	**0**	**8.12**
12	*ndhA*	**1 095**	168	168	0	0	15.34	1 093	32	32	0	0	2.93
13	*ndhC*	363	47	47	0	0	12.95	363	9	9	0	0	2.48

14	*ndhD*	1 512	264	264	0	0	17.46	1 506	79	78	1	1	5.25
15	*ndhE*	306	41	41	0	0	13.40	306	10	10	0	0	3.27
16	*ndhF*	**2 306**	**558**	**513**	**5**	**45**	**24.20**	2 388	171	157	4	14	7.16
17	*ndhG*	531	88	88	0	0	16.57	531	21	21	0	0	3.95
18	*ndhH*	1 182	159	159	0	0	13.45	1 182	47	47	0	0	3.98
19	*ndhI*	513	82	74	1	8	15.98	522	22	22	0	0	4.21
20	*ndhJ*	477	51	51	0	0	10.69	477	12	12	0	0	2.52
21	*ndhK*	684	94	94	0	0	13.74	855	23	23	0	0	2.69
22	*petA*	978	134	134	0	0	13.70	963	37	37	0	0	3.84
23	*petB*	660	56	56	0	0	8.48	648	14	14	0	0	2.16
24	*petD*	483	37	37	0	0	7.66	483	11	10	1	1	2.28
25	*psaA*	2 253	168	168	0	0	7.46	2 253	57	57	0	0	2.53
26	*psaB*	2 205	189	189	0	0	8.57	2 205	42	42	0	0	1.90
27	*psaC*	246	20	20	0	0	8.13	246	7	7	0	0	2.85
28	*psbA*	1 062	83	83	0	0	7.82	1 062	28	28	0	0	2.64
29	*psbB*	1 527	161	161	0	0	10.54	1 527	42	42	0	0	2.75
30	*psbC*	1 422	123	123	0	0	8.65	1 422	26	26	0	0	1.83
31	*psbD*	1 062	62	62	0	0	5.84	1 062	18	18	0	0	1.69
32	*psbE*	252	16	16	0	0	6.35	252	4	4	0	0	1.59
33	*psbH*	222	22	22	0	0	9.91	222	8	8	0	0	3.60
34	*rbcL*	1 428	123	123	0	0	8.61	1 444	51	47	1	4	3.53
35	*rpl14*	369	46	46	0	0	12.47	369	11	11	0	0	2.98
36	*rpl16*							408	20	20	0	0	4.90
37	*rpl20*	405	83	69	1	14	20.49	369	15	15	0	0	4.07
38	*rpl22*	**552**	**163**	**100**	**6**	**63**	**29.53**	**498**	**39**	**35**	**2**	**4**	**7.83**
39	*rpl33*	201	33	33	0	0	16.42	201	8	8	0	0	3.98

40	*rpoA*	1 014	217	206	1	11	21.40	1 014	36	36	0	0	3.55
41	*rpoB*	3 213	413	413	0	0	12.85	3 222	103	103	0	0	3.20
42	*rpoC1*	2 052	282	282	0	0	13.74	2 070	48	47	1	1	2.32
43	*rpoC2*	4 232	837	760	9	77	19.78	4 179	189	186	2	3	4.52
44	*rps11*	417	77	77	0	0	18.47	417	13	10	1	3	3.12
45	*rps12*							**393**	**62**	**62**	**0**	**0**	**15.78**
46	*rps14*	303	37	37	0	0	12.21	303	10	10	0	0	3.30
47	*rps15*	**264**	**72**	**64**	**4**	**8**	**27.27**	**276**	**22**	**22**	**0**	**0**	**7.97**
48	*rps16*							252	18	16	1	2	7.14
49	*rps18*	315	47	47	0	0	14.92	306	4	4	0	0	1.31
50	*rps19*	294	57	57	0	0	19.39	279	20	20	0	0	7.17
51	*rps2*	714	101	101	0	0	14.15	711	18	18	0	0	2.53
52	*rps3*	**654**	**145**	**145**	**0**	**0**	**22.17**	651	37	37	0	0	5.68
53	*rps4*	606	89	89	0	0	14.69	606	22	22	0	0	3.63
54	*rps8*	405	75	75	0	0	18.52	405	15	15	0	0	3.70
55	*ycf1*							**5 712**	**597**	**583**	**7**	**14**	**10.45**
56	*ycf3*	507	43	43	0	0	8.48	507	12	12	0	0	2.37
57	*ycf4*	555	88	88	0	0	15.86	555	25	25	0	0	4.50

3 讨论与结论

3.1 基于叶绿体基因组序列与基于蛋白质编码基因构建的系统发育树冲突

目级水平及牻牛儿苗目内，基于叶绿体基因组序列与基于蛋白质编码基因所构建的系统发育树拓扑结构不一致，其原因可能有两个方面：首先，基因的排列顺序。叶绿体基因组中基因的排列顺序在小的分类单元内保守性较高，变异程度小，随着分类单元的增大，其变异程度往往提高。牻牛儿苗科物种的叶绿体基因组发生过高度的基因重排（Blazier et al.，2011），基因组序列当中的基因顺序在不同物种间差异较大，导致全基因组序列比对时，同一个基因重复出现在基因组的不同位置。第二，基因间隔区的序列变异水平比蛋白质编码基因要高，随着分类单元的扩大，如目级水平，可能由于基因间隔区的序列变异程度太高而导致序列比对困难。因此，已有的研究当中，叶绿体基因组数据应用于目级水平，如石竹目（Liu et al.，2018）、姜目（Wu et al.，

2018）和虎耳草目（Dong et al.，2018）等采用的均为蛋白质编码基因而非全基因组序列。

3.2 桃金娘目及其近缘类群的系统发育关系

桃金娘目和牻牛儿苗目均隶属于蔷薇支，其系统位置一直为研究者所关注（Wang et al.，2009；Soltis et al.，2011）。本研究基于叶绿体基因组数据证实了桃金娘目和牻牛儿苗目的姐妹群关系（BS=100%；PP=1.0）。

使君子科隶属于桃金娘目，形态学及分子系统学证据均不能确认其系统位置，最近的研究结果表明其可能与柳叶菜科+千屈菜科的关系较近（Soltis et al.，2011）。本研究中，基于139 条序列的矩阵（桃金娘目、牻牛儿苗目及外类群的蛋白质编码基因）分析的结果表明，使君子科与柳叶菜科+千屈菜科的关系较近，为姐妹群，获得较高支持率（BS=93%；PP=1.00），进一步，对仅包括桃金娘目的两个数据集（103 条序列的叶绿体基因组与蛋白质编码基因）的分析结果均支持使君子科与柳叶菜科+千屈菜科的姐妹群关系（BS=100%；PP=1.00）。

石榴属在中国植物志中隶属于石榴科，在 *Flora of China* 中被移置于千屈菜科，本研究结果表明石榴属是千屈菜科其余属的姐妹群。菱属在中国植物志及 *Flora of China* 中均隶属于菱科，本研究结果支持将其置于千屈菜科。

3.3 野牡丹科及桃金娘科系统发育分析候选基因

据英文版中国植物志，野牡丹科主要分布于热带及亚热带，有 156~166 属，约 4 500 种，其中我国有 21 属 114 种。已应用于野牡丹科系统发育分析中的叶绿体蛋白质编码基因有 *rbcL*、*ndhF* 和 *rpl16*（Clausing & Renner，2001；Renner，2004）等。桃金娘科有 130 属约 4 500 种，主要分布于地中海、撒哈拉以南非洲、马达加斯加、亚洲热带和温带、澳大利亚、太平洋群岛、美洲热带和南美等地，我国有 10 属 121 种。已应用于桃金娘科系统发育分析中的叶绿体蛋白质编码基因有 *rbcL*、*ndhF*、*rpl16* 和 *matK*（Sytsma et al.，2004；Wilson et al.，2005；Thornhill et al.，2015；Vasconcelos et al.，2017）等。

野牡丹科和桃金娘科二者均为大科，科下包括的属种多。我们的研究结果表明，除上述几个基因外，更多的候选基因可以应用于野牡丹科和桃金娘科的系统发育研究。作为核心 DNA 条形码的 *rbcL* 和 *matK*（CBOL Plant Working Group，2009），序列分化程度较高，在野牡丹科中的变异百分数分别为 8.61%和 27.10%，在桃金娘科中分别为 3.53%和 8.12%。具有"最有潜力的叶绿体 DNA 条形码"之称的 *ycf1* （Dong et al.，2015）在野牡丹科 18 个属中有基因缺失，在桃金娘科中的序列变异百分数为 10.45%。野牡丹科中，序列变异百分数在10%以上的基因有 39 个，5%以上的为 53 个，而在桃金娘科中，序列变异百分数在 10%以上的基因有 2 个，5%以上的为 11 个，这些基因均可成为解决科下系统发育关系的重要候选基因。

参考文献：

BAUSHER MG, SINGH ND, LEE SB, et al.，2006. The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': Organization and phylogenetic relationships to other angiosperms [J]. Bmc Plant Biol, 6(1): 21-31.

BLAZIER JC, GUISINGER MM, JANSEN RK, 2011. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae) [J]. Plant Mol Biol, 76:263-272

CARBONELL-CABALLERO J, ALONSO R, IBANEZ V, et al.，2015. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus* [J]. Mol Biol Evol, 32(8): 2015-2035.

CBOL PLANT WORKING GROUP, 2009. A DNA barcode for land plants. [J]. Proc Natl Acad Sci USA, 106(31): 12794-12797.

CLAUSING G, RENNER SS, 2001. Molecular phylogenetics of Melastomataceae and Memecylaceae: Implications for character evolution [J]. Am J Bot, 88(3): 486-498.

CONTI E, LITT A, WILSON P G, et al.， 1997. Interfamilial relationships in Myrtales: Molecular phylogeny and patterns of morphological evolution [J]. Syst Bot, 22(4): 629-647.

DONG WL, WANG RN, ZHANG NY, et al.， 2018. Molecular evolution of chloroplast genomes of *Orchid* species: insights into phylogenetic relationship and adaptive evolution [J]. Int J Mol Sci, 19(3): 716-735.

DONG WP, XU C, LI CH, et al.， 2015. *ycf1*, the most promising plastid DNA barcode of land plants [J]. Sci Rep, 5: 8348-8352.

DONG WP, XU C, WU P, et al.， 2018. Resolving the systematic positions of enigmatic taxa: Manipulating the chloroplast genome data of Saxifragales[J]. Mol Phylogenet Evol, 126:321-330.

GOLDENBERG R, FRAGA CND, FONTANA AP, et al.， 2012. Taxonomy and phylogeny of *Merianthera* (Melastomataceae) [J]. Taxon, 61(5): 1040-1056.

HILU K W, THOMAS B, KAI M, et al.， 2003. Angiosperm phylogeny based on *matK* sequence information [J]. Am J Bot, 90(12): 1758-1776.

KATOH K, STANDLEY DM, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability [J]. Mol Biol Evol, 30(4): 772-780.

KUMAR S, STECHER G, TAMURA K, 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets [J]. Mol Biol Evol, 33(7): 1870-1874.

LIU X, LI Y, YANG H, et al.， 2018. Chloroplast genome of the folk medicine and vegetable plant *Talinum paniculatum* (Jacq.) Gaertn.: Gene organization, comparative and phylogenetic analysis [J]. Molecules, 23(4): 857-875.

MILLER MA, PFEIFFER W, SCHWARTZ T, 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees[C]. 2010 Gateway Computing Environments Workshop (GCE 2010). New Orleans, LA: 1-8.

PESSOA-FILHO M, MARTINS AM, FERREIRA ME, 2017. Molecular dating of phylogenetic divergence between *Urochloa* species based on complete chloroplast genomes [J]. Bmc Genomics, 18(1): 516-529.

RENNER SS. 1993. Phylogeny and classification of the Melastomataceae and Memecylaceae[J]. Nord J Bot, 13(5):519-540.

RENNER SS. 2004. Bayesian analysis of combined chloroplast loci, using multiple calibrations, supports the recent arrival of Melastomataceae in Africa and Madagascar [J]. Am J Bot, 91(9): 1427-1435.

RONQUIST F, HUELSENBECK JP, 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models [J]. Bioinformatics, 19(12): 1572-1574.

SOLTIS DE, SMITH SA, NICO C, et al.， 2011. Angiosperm phylogeny: 17 genes, 640 taxa [J]. Am J Bot, 98(4): 704-730.

STAMATAKIS A, 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies [J]. Bioinformatics, 30(9): 1312-1313.

SYTSMA K, LITT A, ZJHRA ML, et al.， 2004. Clades, clocks, and continents: Historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the Southern Hemisphere [J]. Int J Plant Sci, 165(S4): S85-S105.

THE ANGIOSPERM PHYLOGENY GROUP, 2016. An update of the Angiosperm Phylogeny

Group classification for the orders and families of flowering plants: APG IV [J]. Bot J Linn Soc, 181(1): 1-20.

THORNHILL AH, HO S YW, KULHEIM C, et al.，2015. Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny [J]. Mol Phylogenet Evol, 93: 29-43.

VASCONCELOS TNC, PROENCA CEB, AHMAD B, et al.，2017. Myrteae phylogeny, calibration, biogeography and diversification patterns: Increased understanding in the most species rich tribe of Myrtaceae [J]. Mol Phylogenet Evol, 109: 113-137.

WANG HC, MOORE MJ, SOLTIS PS, et al.，2009. Rosid radiation and the rapid rise of angiosperm-dominated forests [J]. Proc Natl Acad Sci USA, 106(10): 3853-3858.

WILSON PG, O'BRIEN MM, HESLEWOOD MM, et al.，2005. Relationships within Myrtaceae sensu lato based on a *matK* phylogeny [J]. Plant Syst Evol, 251(1): 3-19.

WU ML, LI Q, XU J, et al.，2018. Complete chloroplast genome of the medicinal plant *Amomum compactum*: Gene organization, comparative analysis, and phylogenetic relationships within Zingiberales [J]. Chin Med, 13(1): 10-21.

WU Y, LIU F, YANG DG, et al.，2018. Comparative chloroplast genomics of *Gossypium* species: Insights into repeat sequence variations and phylogeny [J]. Front Plant Sci, 9: 376-407.

ZHANG YJ, LI DZ, 2011. Advances in phylogenomics based on complete chloroplast genomes[J]. Plant Divers Resour, 33(4): 365-375. [张韵洁，李德铢，2011. 叶绿体系统发育基因组学的研究进展 [J]. 植物分类与资源学报，33（4）：365–375.]

ZHU XY, CHASE MW, QIU YL, et al.，2007. Mitochondrial *matR* sequences help to resolve deep phylogenetic relationships in rosids [J]. Bmc Evol Biol, 7(1): 217-231.